

J. Clin. Chem. Clin. Biochem.
Vol. 21, 1983, pp. 373–379

The First External Quality Control Survey (EQCS) for Antithyroglobulin- and Antimicrosomal Antibody Determination

By R. Gärtner, K. Horn, A. König

Medizinische Klinik Innenstadt der Universität München West-Germany and

H. Schatz

Medizinische Klinik und Poliklinik der Universität Gießen, West-Germany

(Received March 3, 1982)

Summary: Fourteen laboratories participated in an external quality control survey (EQCS) for antithyroglobulin- and antimicrosomal-antibody determination. Seventeen frozen serum samples from individuals suffering from various thyroid diseases were dispatched by mail. Different test systems were used by the participants including the haemagglutination test, radioimmunoassay (RIA), enzyme immunoassay (ELISA) and immunofluorescence test. The different test results of each participant were sized and could be compared by giving place numbers to each test result. Inter-laboratory quality control was performed by evaluation of a rank correlation coefficient. There was close agreement between the haemagglutination test results of antimicrosomal antibody determination with different test kits, and these tests showed good precision. Antithyroglobulin antibody titers from the two haemagglutination test-kits used, however, were not directly comparable. When sized by place numbers, antithyroglobulin antibody results from haemagglutination test, RIA and ELISA were in close agreement. The least correspondence was shown by the immunofluorescence test, in the antithyroglobulin antibody determination.

This EQCS demonstrates the necessity of international standardisation and the use of test systems which allow quantitative determination of antithyroglobulin- and antimicrosomal-antibody.

Erster Ringversuch zur Qualitätskontrolle der Bestimmung von Thyreoglobulin- und mikrosomalen Antikörpern

Zusammenfassung: An einem ersten Ringversuch zur Qualitätskontrolle der Bestimmung von Thyreoglobulin- und mikrosomalen Antikörpern im Serum nahmen 14 Laboratorien teil. Siebzehn Serumproben von Patienten mit definierten Schilddrüsenerkrankungen wurden verschickt. Die hauptsächlich im Handel befindlichen Testsysteme (Hämagglutinationsteste, Radioimmunoassays, Enzymimmunoassays, Immunfluoreszenz-Teste) und nur wenige eigen entwickelte wurden verwendet. Da die Testsysteme semiquantitative bzw. qualitative Ergebnisse lieferten und deshalb nicht direkt vergleichbar waren, wurden die Testergebnisse der einzelnen Laboratorien nach Platzziffern geordnet. Zur Präzisionskontrolle der einzelnen Testkits wurde ein Rang-Korrelationskoeffizient errechnet.

Die beste Übereinstimmung und Präzision erzielten die Hämagglutinationsteste zur Bestimmung der mikrosomalen Antikörper. Bei der Bestimmung der Thyreoglobulin-Antikörper waren die Titer der zwei verschiedenen, kommerziell erhältlichen Hämagglutinationsteste unterschiedlich und nicht direkt vergleichbar. Die radioimmunologischen und enzymimmunologischen Testverfahren waren nach Verteilung von Platzziffern für die einzelnen Serumproben größtmäßig vergleichbar mit den Hämagglutinationstesten. Die Präzision der Thyreoglobulin-Antikörper-Test-Kits war insgesamt schlechter als die der für mikrosomale Antikörper. Die größten Schwankungen zeigten die verschiedenen Immunfluoreszenztest-Kits. Diese Untersuchung zeigt die Notwendigkeit, die einzelnen Testsysteme mehr zu standardisieren und Bestimmungsmethoden zu entwickeln, die eine bessere Quantifizierung der Ergebnisse ermöglichen.

Introduction

Autoantibodies against various components of the thyroid gland have gained in importance, especially with reference to the aetiology (1–4) and diagnosis (5, 6) of various diseases of the thyroid gland. This increasing importance is due not least to more sensitive test systems, which were developed during the last few years. The following applications of autoantibody determination have attracted special attention:

- 1) Differentiation between *Graves'* disease and autonomous adenomata and multifocal autonomy, respectively;
- 2) follow-up of *Graves'* disease under medical treatment;
- 3) investigation of the role that autoantibodies may play in the differentiation between immunogenic thyroid disorders and simple endemic goiter (7);
- 4) investigation of the natural history of *Hashimoto's* disease with special regard to the development of hypothyroidism (8).

Basically four different systems are available for the determination of antithyroglobulin- and antimicrobial-antibody: haemagglutination tests, radio-ligand-assays, enzyme immunoassays and finally immunofluorescence tests. The aim of this first external quality control survey was to determine whether the results of these different determinations are comparable. The survey was initiated by the German Endocrine Society, Thyroid Section. Fourteen German laboratories took part, some of them using more than one of the methods cited above.

Materials and Methods

Samples

Serum samples were drawn from 17 different patients with various diseases of the thyroid (tab. 1). It should be noted that the diagnoses were made according to clinical and functional criteria only, i.e. none of the antibodies to be tested underlay the diagnosis given in table 1, except in *Hashimoto's* disease (serum no. 3, 6, 16). The frozen serum samples were dispatched by mail, and the laboratories participating only knew the numbers of the serum samples.

Evaluation

Determination of antithyroglobulin antibodies

The following test systems were used:

- 1) Haemagglutination test (Fujizoki distributed by Mast): 6 participants.
- 2) Haemagglutination test (Wellcome): 5 participants.
- 3) Radioligand assay (CIS/IDW): 3 participants.

- 4) Radioligand assay (Serono): 2 participants.

- 5) Immunofluorescence test (Mast, Proma, Travenol, own method): each with 1 participant.

- 6) Enzyme immunoassay (own method): 1 participant.

In view of the different methodologies (and the fact that the results were qualitative, semiquantitative or quantitative) a standardized evaluation of the data was extremely difficult. Therefore we evaluated the results obtained by immunofluorescence tests separately from those obtained by haemagglutination tests. In addition, we tried to obtain an overall evaluation as follows: The different test results of each participant were sized and a place number was given to each test result. Furthermore a rank correlation analysis (according to *Spearman* (9)) was applied to the ranks obtained from the test results for evaluation of inter-laboratory quality control.

Distribution of place numbers

The results obtained by each of the participants using their special method were sized and numbered; i.e. place number 1 was given to the highest values of antithyroglobulin antibody and so forth. If the same result was reported more than once, the appropriate place number was given to each of these results, omitting the following place numbers. An example: Participant no. 2 obtained the highest antithyroglobulin antibody titer (1:25600) from sample no. 5 and 6; these received place number 1, place no. 2 was omitted. The next lower titer (1:6400) was obtained from sample no. 3, which received place no. 3 and so forth. The negative results were evaluated separately: To each negative result place no. 17 was given.

Distribution of ranks (used in rank correlation analysis)

Basically the distribution of ranks was done as described above. The treatment of ties, however, was different: The rank given to equal results was computed as the mean value of the ranks given to each of the equal results. In the example above the rank 1.5 ($(1+2)/2$) was given to the two highest results. The same method was also applied to negative results. If, for example, four negative results were reported, to each of these results rank 15.5 ($(12+13+14+15)/4$) was given.

Tab. 1. Characterization of samples.

Sample no.	Diagnosis*
1	<i>Graves'</i> disease, euthyroid
2	<i>Graves'</i> disease, euthyroid for 20 years
3	<i>Hashimoto's</i> disease
4	Thyrotoxicosis without eye signs
5	Thyrotoxicosis without eye signs
6	<i>Hashimoto's</i> disease, euthyroid
7	<i>Graves'</i> disease
8	Subacute thyroiditis
9	<i>Graves'</i> disease, euthyroid
10	<i>Graves'</i> disease
11	Euthyroid <i>Graves'</i> disease
12	<i>Graves'</i> disease
13	<i>Graves'</i> disease, euthyroid
14	Sample 13+100 µg/l thyroglobulin standard
15	<i>Graves'</i> disease
16	<i>Hashimoto's</i> disease
17	Sample 13+300 µg/l thyroglobulin standard

* Diagnoses were received from clinical and laboratory findings, antibody titers were not used for diagnosis except in *Hashimoto's* thyroiditis.

$(14+15+16+17)/4$) was given. Theoretically if 17 equal results had been obtained from one laboratory, rank 9 would have been given to each of them. A rank correlation analysis (9) was done with the 17 results of each participant. The reference quantities R_i were computed using the following formula:

$$R_i = (r_{i1} + r_{i2} + \dots + r_{i17})/17.$$

r_{ij} : rank given to results of participant j for sample i ;

i.e. the mean rank of participants for a particular sample was taken as reference quantity. The rank correlation coefficient r_j^+ of participant j was computed by

$$r_j^+ = 1 - \frac{6 \sum_{i=1}^{17} (r_{ij} - R_i)^2}{17(17^2 - 1)}$$

In this way, an inter-laboratory comparison of the results was possible.

Results obtained by the haemagglutination test were evaluated separately from those obtained by the immunofluorescence test. Furthermore rank correlation analysis was done as described above.

Results

Antithyroglobulin antibodies

Haemagglutination test

The absolute antibody titers obtained by the two different test kits (Fujizoki, Wellcome) could not be

compared. We therefore set the highest titer obtained by the Fujizoki (1:409 600) equal to the highest titer obtained by the Wellcome kit (1:5120). The next lower titers were treated in the same way and so on. As one may see in figure 1 the results obtained by the two different test kits can be compared by this method fairly well. There is, however, one remarkable difference: The samples with very low titers are assessed as negative more often by the Wellcome kit than by the kit from Fujizoki. The results obtained by haemagglutination test did not depend on the two thyroglobulin concentrations (100 µg/l, 300 µg/l), which were added to the samples 13 to produce samples 14 and 17.

Immunofluorescence test

The immunofluorescence test naturally gives less differentiation. Figure 2 shows that 14 out of the 17 samples were determined either positive or negative by the different participants. The results of sample 13 and 17 demonstrate the dependence on the high thyroglobulin concentration of the heterologous immunofluorescence tests. The concentration antithyroglobulin antibody in sample 17 is assessed as negative by two participants, whereas sample 13 is assessed as positive by all four participants.

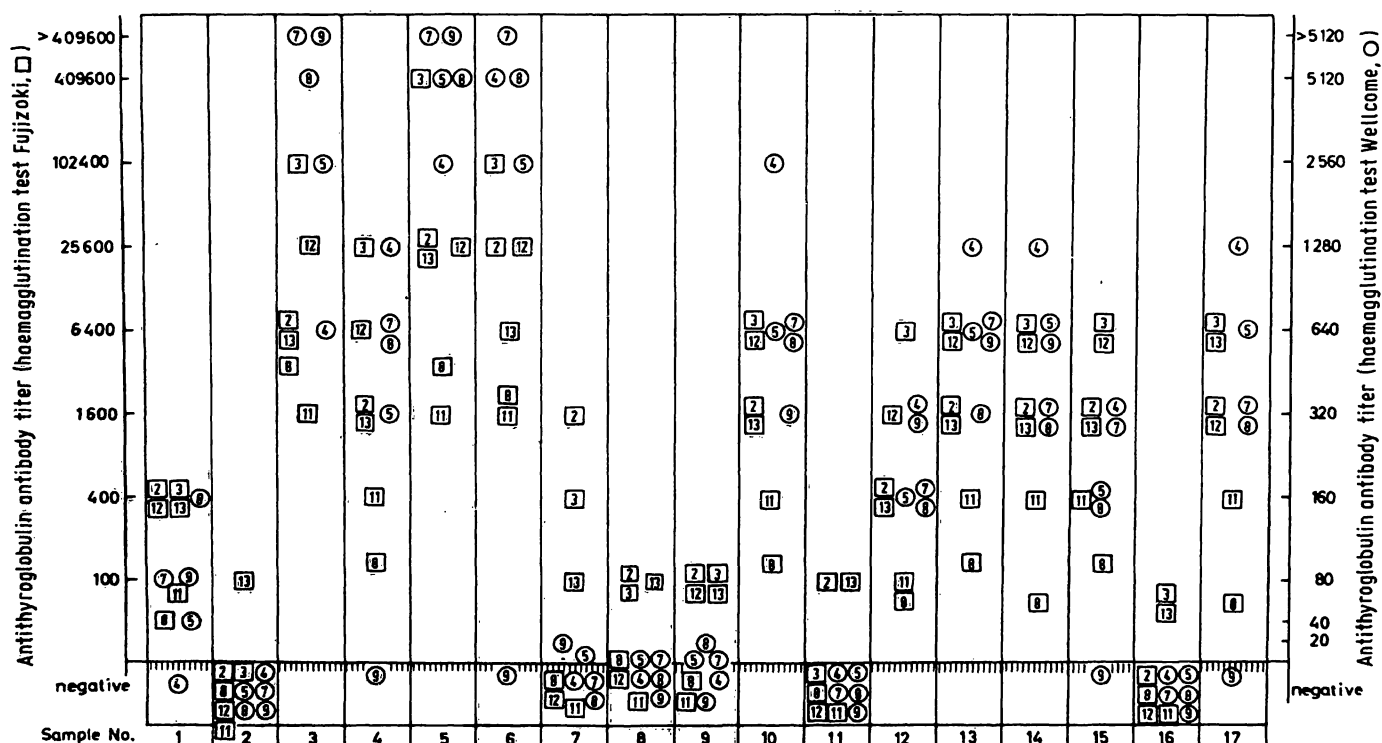


Fig. 1. Antithyroglobulin antibody titers obtained by haemagglutination test of all serum samples, titers from Fujizoki (\square , $n = 6$) are given on the left, titers from Wellcome (\circ , $n = 5$) on the right ordinate. Numbers inside the symbols represent the numbers given to the different participants.

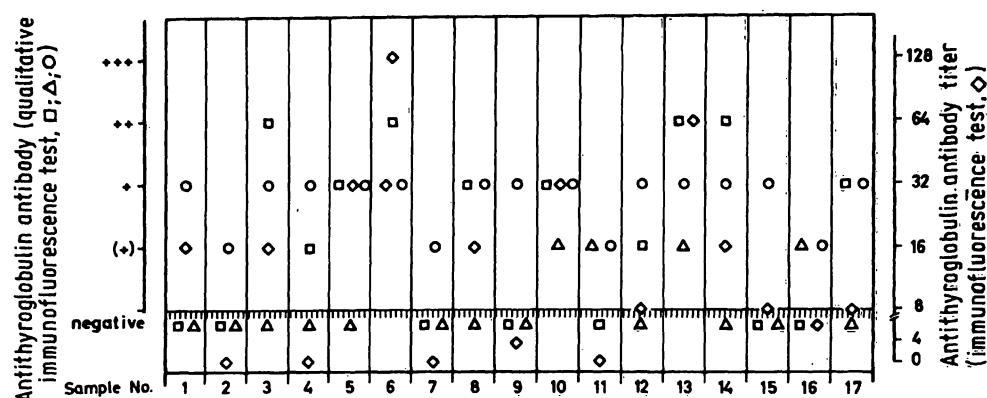


Fig. 2. Antithyroglobulin antibody results obtained from immunofluorescence test. Qualitative values from Mast, Proma and the own levels are given on the left, semiquantitative values from Travenol on the right ordinate.

□ Proma (8) △ Mast (8) ◇ Travenol (10) ○ Own method (15)

Evaluation of the place numbers

Place numbers were distributed to each result regardless of whether it was determined by RIA, ELISA, immunofluorescence test or haemagglutination test (see above). Figure 3 shows that the majority of

the participants yielded identical results, except for sample Nos. 7, 8 and 9. The results of these three samples show the greatest differences: whereas about half the participants assessed these samples as negative, the rest assessed these samples as positive, even markedly positive.

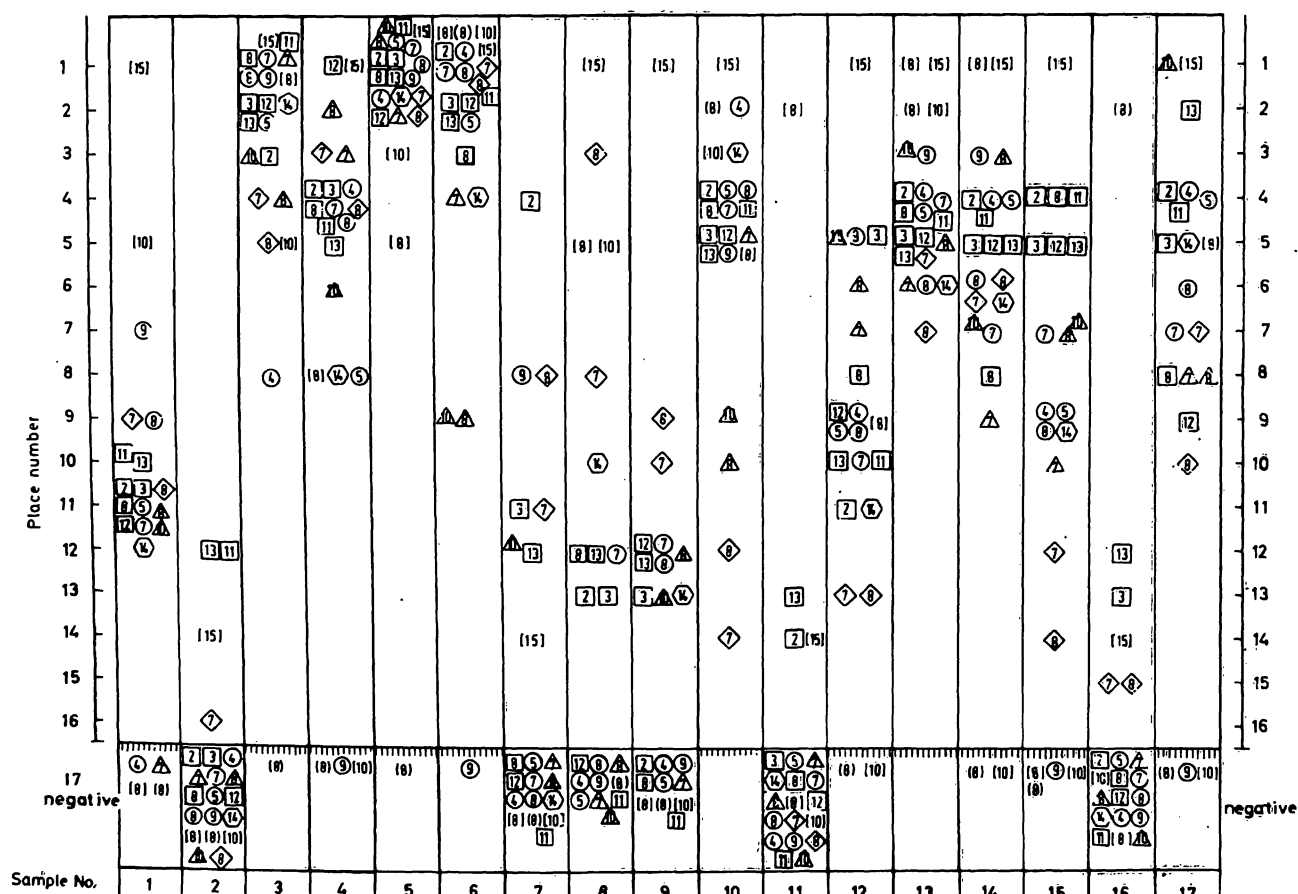


Fig. 3. Diagram of place numbers from antithyroglobulin antibody values. The numbers inside the symbols represent the number given to the different participants.

□ Fujizoki ○ Wellcome △ CIS-IDW ◇ Sersono ○ Own method ELISA,
immunofluorescence test [] Travenol, Proma, Own method () Mast

Rank correlation analysis (Spearman)

High rank correlation coefficients are yielded by the results of the haemagglutination tests, except for participant no. 9. A slightly lower rank correlation coefficient was yielded by radioimmunological methods, immunofluorescence tests giving the lowest (tab. 2).

Antimicrosomal antibodies

Haemagglutination test

The titers obtained by haemagglutination test Wellcome are directly comparable to those obtained by haemagglutination test Fujizoki. The coincidence is fairly good (fig. 4), except for the results from samples 6, 8 and 9 which were assessed as markedly positive by haemagglutination test Fujizoki, whereas by haemagglutination test Wellcome the results were more often assessed as negative.

Immunofluorescence test

The results obtained by immunofluorescence test coincide fairly well (except for two samples which are assessed as negative as well as positive). The correspondence of these results was markedly better, compared with the immunofluorescence test results of the antithyroglobulin antibodies (fig. 5).

Tab. 2. Rank correlation coefficient for antithyroglobulin antibody (all participants). If $r^{\dagger} \leq 0.564$ there is no linear correlation between the ranks of participant no. i and the mean values of all ranks (R_j , see formula no. 1).

Test systems	Participant no.	Rank correlation coefficient r^{\dagger}
<i>Haemagglutination test</i>		
Fujizoki	2	0.896
	3	0.964
	8	0.949
	11	0.954
	12	0.962
Wellcome	13	0.925
	4	0.889
	5	0.965
	8	0.966
	9	0.577
Radioimmunoassay	7	0.930
	8	0.879
	10	0.818
	8	0.737
	7	0.860
Own method (ELISA)	10	0.737
	14	0.928
Immunofluorescence test	8	0.795
	8	0.203
	10	0.713
	15	0.773

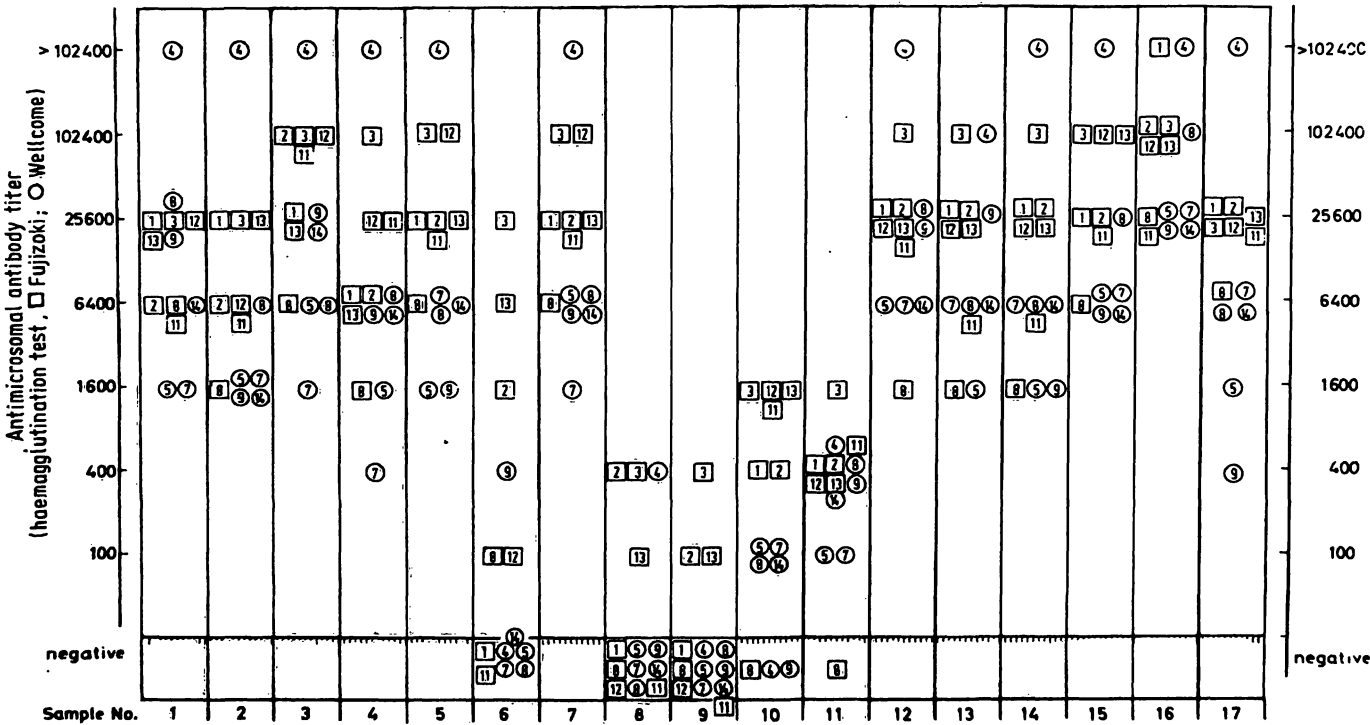


Fig. 4. Diagram of antimicrosomal antibody titers from haemagglutination test (Fujizoki, Wellcome). Numbers inside the symbols represent the numbers given to the different participants.

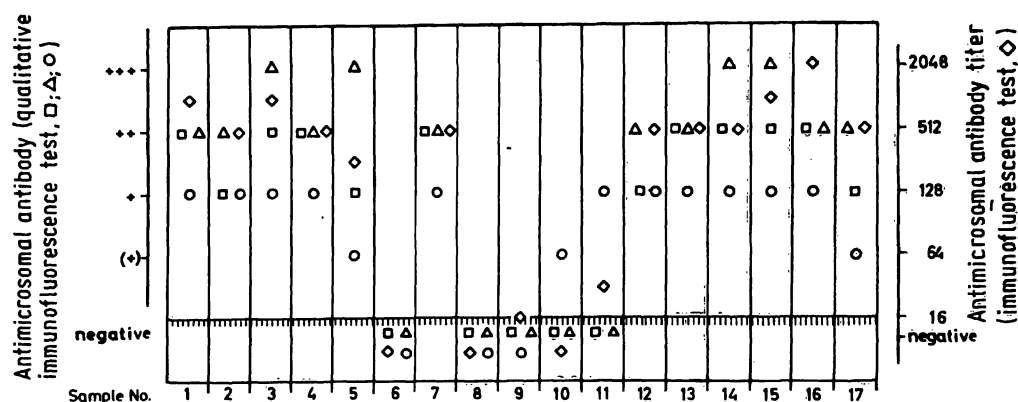


Fig. 5. Diagram of the results obtained by immunofluorescence test for antimicrosomal antibody. Qualitative values (Proma, Mast, own method) are given on the left, semiquantitative values (Travenol) on the right ordinate.

□ Proma (8) △ Mast (8) ◇ Travenol (10) ○ Own method (15)

Rank correlation analysis

The rank correlation coefficients (tab. 3) of the participants using haemagglutination test are all markedly high. Though the coefficients of those participants using immunofluorescence test are lower they are still significant on the 1% level.

Tab. 3. Rank correlation coefficient for antimicrosomal antibody (all participants).

Test systems	Participant no.	Rank correlation coefficient r^+
<i>Haemagglutination test</i>		
Fujizoki	1	0.887
	2	0.898
	3	0.866
	8	0.884
	11	0.846
	12	0.914
	13	0.908
Wellcome	4	0.839
	5	0.962
	7	0.803
	8	0.899
	9	0.845
	14	0.918
<i>Immunofluorescence test</i>		
Proma	8	0.813
Mast	8	0.761
Travenol	10	0.898
Own method	15	0.714

Discussion

The inter-laboratory test results of the antimicrosomal antibody determinations were in good correspondence; the titers yielded by the two haemagglutination test kits used can be compared directly. Furthermore the results obtained by immunofluorescence test show good agreement with those obtained by haemagglutination test.

The results of antithyroglobulin antibodies show greater differences. The absolute antibody titers obtained by the two different haemagglutination test kits (Fujizoki, Wellcome) could not be compared directly; however, they were made comparable by setting the highest titer (1:409600) of the Fujizoki assay equal to the highest values (1:5120) of the Wellcome assay and so on for the next lower titers. Whereas haemagglutination test, RIA and ELISA were not disturbed by thyroglobulin added to sample no. 13, the heterologous immunofluorescence test kits (Mast, Proma, Travenol) gave lower or negative results when thyroglobulin was added (sample 14, 17).

The rank correlation coefficient was best in haemagglutination test kits; worse but still significant coefficients were obtained with immunofluorescence test kits. The use of the rank correlation analysis, however, needs some further explanations:

1) The method used by most of the participants is haemagglutination test (11 participants), whereas 7 participants used radioimmunological methods and

only 4 immunofluorescence tests. It is obvious that the mean value of the ranks distributed to each of the results is biased by the method with the highest number of participants (i.e. haemagglutination test). Therefore the very high rank correlation coefficients of these participants are mainly due to the fact that they belong to the strongest group, because the mean value of ranks served as the reference quantity in the correlation analysis.

2) Rank correlation coefficients can only serve as a rough method to check the inter-laboratory comparability or deviation. It would be wrong to conclude that the results yielded by one participant are correct just because they show a high correlation coefficient. Furthermore, in external quality control surveys for assays without a reference method the mean value of the majority of the participants is not necessarily the correct result of the parameter under investigation.

It seems to be important to us that there is no interference of antithyroglobulin antibody in antimicrobial antibody assays and vice versa. This is demonstrated by samples no. 2, 7, 11, 16 with high positive antimicrobial antibody and negative antithyroglobulin antibody titers and sample no. 6 with high antithyroglobulin antibody titer and low (or even negative) antithyroglobulin antibody titer respectively.

In spite of these encouraging findings there is still room for improvements, especially in antithyroglobulin antibody test kits. It is desirable to achieve a better standardization in order to be able to compare the results obtained by different test kits. This would be possible, if RIA or ELISA for antithyroglobulin antibody (10) and antimicrobial antibody (11) were used more widely, and international standards were available to calibrate the different assay systems.

References

1. Doniach, D. (1975) *Clin. Endocrinol. Metab.* 4, 267–285.
2. Volpé, R. (1977) *Ann. Intern. Med.* 87, 86–99.
3. Volpé, R. (1978) *Clinics Endocrinol. Metab.* 7, 3–29.
4. Roitt, J. M., Doniach, D., Campbell, P. N. & Hudson, R. V. (1956) *Lancet II*, 820.
5. Schatz, H., Kühn, A., Zimmermann, U., Teuber, J., Greve, S. & Federlin, K. (1981) *Aktuelle Endokrinol.* 2, 161–167.
6. Schmolke, B. (1981) *Lab. Med.* 5, 173–182.
7. Doniach, D., Bottazzo, G. F. & Russell, R. C. G. (1979) *Clinics Endocrinol. Metab.* 8, 63–80.
8. Hedley, A. J., Thjodleifsson, B., Donald, D., Beck, J. S., Crooks, J., Chester, M. I. & Hall, R. (1977) *Clin. Endocrinol.* 7, 377.
9. Spearman, C. (1904) *Amer. J. Psychol.* 15, 72–101.
10. Bodlaender, P., Arjonilla, J. R. & Twoney, S. L. (1978) *Clin. Chem.* 24, 272–274.
11. Kung, V. T., Weber, P. M. & dos Remedios, L. V. (1981) *Clin. Chem.* 27, 39–42.

Dr. med. R. Gärtner
Med. Klinik Innenstadt
Ziemssenstraße 1
D-8000 München 2

